BANQUE COMMUNE D'EPREUVES

CODE EPREUVE :
283
CCIP_M2_S

Concepteur : H.E.C. – E.S.C.P. – E.A.P.

OPTION SCIENTIFIQUE

MATHEMATIQUES II

Mercredi 7 mai 2008, de 14 h. à 18 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.
Ils ne doivent faire usage d'aucun document : **l'utilisation de toute calculatrice et de tout matériel électronique est interdite.**
Seule l'utilisation d'une règle graduée est autorisée.

Toutes les variables aléatoires qui interviennent dans ce problème sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$. Sous réserve d'existence, on note $E(X)$ et $V(X)$ respectivement, l'espérance et la variance d'une variable aléatoire réelle X quelconque. Pour toute variable aléatoire réelle X admettant une densité sur $\mathbb{R}$, notée f_X, on note $\mathcal{D}_X$ l'ensemble des réels s pour lesquels la variable aléatoire e^{sX} admet une espérance, et on note Φ_X la fonction définie sur $\mathcal{D}_X$ par : $\Phi_X(s) = E(e^{sX})$.

On **admet** les résultats suivants :

- si deux variables aléatoires X et Y sont telles que Φ_X et Φ_Y coïncident sur un intervalle ouvert non vide, alors X et Y ont la même loi ;
- si n est un entier naturel non nul, et $X_1, X_2, \ldots, X_n$ des variables aléatoires réelles quelconques, mutuellement indépendantes, alors, pour tout entier p de $[\![1, n-1]\!]$ et pour toutes fonctions réelles continues φ_1 et φ_2, les variables aléatoires $\varphi_1(X_1, \ldots, X_p)$ et $\varphi_2(X_{p+1}, \ldots, X_n)$ sont indépendantes ;
- si X et Y sont des variables aléatoires indépendantes admettant une espérance, alors XY admet une espérance, et $E(XY) = E(X)E(Y)$.

La fonction exponentielle est également notée exp. On rappelle que : $\displaystyle\int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2}\right) dx = \sqrt{2\pi}$.

Dans tout le problème, U désigne une variable aléatoire qui suit la loi normale centrée réduite.

Préliminaire

On rappelle que, pour tout s de $\mathcal{D}_X$, on a : $\displaystyle\Phi_X(s) = \int_{-\infty}^{+\infty} \exp(sx) f_X(x) dx$.

1. Soit a un réel non nul et b un réel quelconque.

a) Montrer que l'intégrale $\displaystyle\int_{-\infty}^{+\infty} \exp(-ax^2) dx$ est convergente si et seulement si $a > 0$, et vaut alors $\sqrt{\dfrac{\pi}{a}}$.

b) En déduire que l'intégrale $\int_{-\infty}^{+\infty} \exp(-ax^2+bx)dx$ est convergente si et seulement si $a>0$, puis montrer que dans ces conditions, on a : $\int_{-\infty}^{+\infty} \exp(-ax^2+bx)dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a}\right)$.

2. a) Déterminer $\mathcal{D}_U$; pour tout s de $\mathcal{D}_U$, calculer $\Phi_U(s)$.

b) On pose : $Z = U^2$. Établir que : $\mathcal{D}_Z =]-\infty, \frac{1}{2}[$; montrer, à l'aide du théorème de transfert, que pour tout réel s de $\mathcal{D}_Z$, on a : $\Phi_Z(s) = (1-2s)^{-1/2}$.

3. Soit X une variable aléatoire réelle à densité, et soit μ et β deux réels quelconques.

a) Montrer qu'un réel s appartient à $\mathcal{D}_{\mu X+\beta}$ si et seulement si μs appartient à $\mathcal{D}_X$, et que dans ce cas, on a : $\Phi_{\mu X+\beta}(s) = \exp(\beta s)\Phi_X(\mu s)$.

b) On suppose que X suit une loi γ de paramètre ν, où ν est un réel strictement positif.

Montrer que : $\mathcal{D}_X =]-\infty, 1[$; pour tout s de $\mathcal{D}_X$, établir la formule : $\Phi_X(s) = (1-s)^{-\nu}$. De même, déterminer $\mathcal{D}_{2X}$; pour tout s de $\mathcal{D}_{2X}$, calculer $\Phi_{2X}(s)$.

Partie I. Loi du χ^2 centré

Soit r un entier supérieur ou égal à 1. On considère une variable aléatoire X_r suivant la loi Γ de paramètres $\left(2, \frac{r}{2}\right)$, c'est-à-dire que X_r possède une densité f_{X_r} donnée par :

$$f_{X_r}(x) = \begin{cases} \dfrac{1}{2^{\frac{r}{2}} \times \Gamma\left(\frac{r}{2}\right)} \times x^{\frac{r}{2}-1} \times \exp\left(-\dfrac{x}{2}\right) & \text{si } x > 0 \\ 0 & \text{si } x \leqslant 0 \end{cases}$$

On dit que X_r suit une loi du χ^2 (« chi deux ») centré à r degrés de liberté, et on note : $X_r \hookrightarrow \chi^2(r)$.

1. Étudier les variations de f_{X_4} et tracer sa courbe représentative dans un repère orthogonal du plan.

2. a) Montrer que la variable aléatoire $\frac{X_r}{2}$ suit une loi γ de paramètre $\frac{r}{2}$. En déduire $E(X_r)$ et $V(X_r)$.

b) Déterminer $\mathcal{D}_{X_r}$; pour tout s de $\mathcal{D}_{X_r}$, calculer $\Phi_{X_r}(s)$.

3. Soit n un entier de $\mathbb{N}^*$. On considère n variables aléatoires indépendantes $U_1, U_2, \ldots, U_n$ de même loi que U. Pour tout i de $[\![1, n]\!]$, on pose : $Z_i = U_i^2$.

a) Vérifier que X_1 et U^2 sont de même loi.

b) On pose : $W_n = \sum_{i=1}^{n} Z_i$. Quelle est la loi de probabilité de W_n ?

c) Déterminer $\mathcal{D}_{W_n}$, et pour tout s de $\mathcal{D}_{W_n}$, exprimer $\Phi_{W_n}(s)$ en fonction de s et de n. Établir une relation entre $\Phi_{W_n}(s)$ et $\Phi_{Z_1}(s), \Phi_{Z_2}(s), \ldots, \Phi_{Z_n}(s)$.

4. Soit T une variable aléatoire qui suit la loi normale centrée, de variance σ^2 inconnue, σ étant un réel strictement positif. Pour n entier supérieur ou égal à 2, on dispose d'un n-échantillon indépendant, identiquement distribué (i.i.d.), $T_1, T_2, \ldots, T_n$ de la loi de T. On considère la variable aléatoire S_n définie par : $S_n = \frac{1}{n}\sum_{i=1}^{n} T_i^2$.

a) Montrer que S_n est un estimateur sans biais et convergent du paramètre σ^2.

b) Soit α un réel vérifiant : $0 < \alpha < 1$, et soit k_α le réel strictement positif tel que : $P([W_n \geqslant k_\alpha]) = 1 - \alpha$. Montrer que l'intervalle $\left]0, \frac{nS_n}{k_\alpha}\right]$ est un intervalle de confiance de σ^2 au risque α.

Partie II. Loi du χ^2 décentré

On considère une suite $(M_j)_{j\geqslant 1}$ de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, mutuellement indépendantes, telles que pour tout j de $\mathbb{N}^*$, M_j suive la loi normale d'espérance m_j ($m_j \in \mathbb{R}$) et de variance égale à 1.

Pour n entier de $\mathbb{N}^*$, on pose : $Y_n = \sum_{j=1}^{n} M_j^2$ et $\lambda_n = \sum_{j=1}^{n} m_j^2$.

On dit que Y_n suit une loi du χ^2 décentré à n degrés de liberté, de paramètre de décentrage λ_n, et on note : $Y_n \hookrightarrow \chi^2(n, \lambda_n)$.

1. Dans cette question *uniquement*, on suppose que l'entier n est égal à 1.

a) Montrer les deux égalités suivantes : $E(U^3) = 0$ et $E(U^4) = 3$.

b) En déduire, en fonction de λ_1, les valeurs respectives de $E(Y_1)$ et de $V(Y_1)$.

c) Montrer que : $\mathcal{D}_{Y_1} =]-\infty, \frac{1}{2}[$ et établir, pour tout réel s de $\mathcal{D}_{Y_1}$, la formule suivante :

$$\Phi_{Y_1}(s) = (1-2s)^{-1/2} \times \exp\left(\frac{s\lambda_1}{1-2s}\right)$$

2. Soit n un entier de $\mathbb{N}^*$.

a) Calculer $E(Y_n)$ et $V(Y_n)$ en fonction de n et λ_n.

b) On admet que l'on a : $\mathcal{D}_{Y_n} =]-\infty, \frac{1}{2}[$. Pour tout s de $\mathcal{D}_{Y_n}$, exprimer $\Phi_{Y_n}(s)$ en fonction de s, n et λ_n.

Partie III. Nombre aléatoire de degrés de liberté

Sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, on considère une variable aléatoire X à valeurs dans $\mathbb{R}$ admettant une espérance $E(X)$, et une variable aléatoire K à valeurs dans $\mathbb{N}$. On note N_K l'ensemble des entiers naturels k vérifiant $P([K=k]) > 0$, et on suppose que pour tout entier k de N_K, la variable aléatoire X admet une espérance pour la probabilité conditionnelle $P_{[K=k]}$, notée $E(X/K=k)$.

On admet alors l'égalité suivante : $(\star)$ $E(X) = \sum_{k\in N_K} E(X/K=k)P([K=k])$

Soit g l'application définie sur $\mathbb{N}$ par : $g(k) = \begin{cases} E(X/K=k) & \text{si } k \in N_K \\ 0 & \text{sinon} \end{cases}$.

1. *Vérification de la formule $(\star)$ sur un exemple.*

Soit $(J_i)_{i\geqslant 1}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$, indépendantes et de même loi uniforme sur l'intervalle $[0,1]$. Pour tout k de $\mathbb{N}^*$, on pose : $X_k = \sup_{1\leqslant i\leqslant k}(J_i)$; autrement dit, pour tout ω de Ω, $X_k(\omega) = \max_{1\leqslant i\leqslant k} J_i(\omega)$. Soit K une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$ qui suit la loi uniforme discrète sur $[\![1, n]\!]$ ($n \in \mathbb{N}^*$). On suppose que K est indépendante des variables aléatoires de la suite $(J_i)_{i\geqslant 1}$.

Pour tout ω de Ω, on pose : $X(\omega) = \max_{1\leqslant i\leqslant K(\omega)} J_i(\omega)$, et on admet que X est une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$.

a) Établir, pour tout entier k de $[\![1, n]\!]$ et pour tout réel x, la relation : $P_{[K=k]}([X \leqslant x]) = P([X_k \leqslant x])$.

b) Déterminer la fonction de répartition F_X de la variable aléatoire X.

c) En déduire que X est une variable aléatoire à densité, qui admet une espérance $E(X)$ que l'on exprimera en fonction de $\sum_{k=1}^{n} \frac{k}{k+1}$.

d) Vérifier l'égalité $(\star)$: $E(X) = E(g(K))$.

2. Soit $(U_i)_{i\geqslant 1}$ une suite de variables aléatoires indépendantes qui suivent la loi normale centrée réduite. Soit K une variable aléatoire indépendante des variables aléatoires de la suite $(U_i)_{i\geqslant 1}$, qui suit la loi de Poisson de paramètre $\frac{\lambda}{2}$ strictement positif.

Pour n entier de $\mathbb{N}^*$, on pose : $H_n = U_1^2 + U_2^2 + \cdots + U_{n+2K}^2$. On admet que H_n est une variable aléatoire à densité à valeurs positives, et que $\mathcal{D}_{H_n} =]-\infty, \frac{1}{2}[$.
Soit s un réel de $]-\infty, \frac{1}{2}[$.

a) Montrer que pour tout k de $\mathbb{N}$, la loi conditionnelle de H_n sachant $[K = k]$ est la loi de la variable aléatoire W_{n+2k} définie dans la question I.3.b.

b) En posant : $X = e^{sH_n}$, déterminer, pour tout entier k de $\mathbb{N}$, l'expression de $g(k)$ en fonction de k.

c) Établir la formule suivante :

$$E(g(K)) = (1-2s)^{-n/2} \times \exp\left(\frac{\lambda s}{1-2s}\right)$$

d) En utilisant l'égalité $(\star)$, admise au début de cette partie, avec $X = e^{sH_n}$, déterminer la loi de H_n.

e) À l'aide de la question III.2.a, montrer que pour tout entier n supérieur ou égal à 3, on a :

$$E\left(\frac{1}{H_n}\right) = E\left(\frac{1}{n-2+2K}\right)$$

Partie IV. Estimateur de James-Stein

Soit p un entier supérieur ou égal à 3. On suppose qu'un modèle aléatoire défini sur $(\Omega, \mathcal{A}, P)$ comporte p paramètres réels inconnus $\theta_1, \ldots, \theta_p$ non tous nuls. Un échantillon d'observations statistiques permet d'exhiber des estimateurs $\widehat{\theta_1}, \ldots, \widehat{\theta_p}$ *sans biais* des paramètres $\theta_1, \ldots, \theta_p$ respectivement. On suppose que les variables aléatoires $\widehat{\theta_1}, \ldots, \widehat{\theta_p}$ sont indépendantes et suivent une loi normale de variance égale à 1.

On pose : $\theta = (\theta_1, \ldots, \theta_p)$, $\widehat{\theta} = (\widehat{\theta_1}, \ldots, \widehat{\theta_p})$, $B_p = \sum_{j=1}^{p} \widehat{\theta_j}^2$ et $b_p = \sum_{j=1}^{p} \theta_j{}^2$.

On dit que *le vecteur aléatoire $\widehat{\theta}$ est un estimateur sans biais du paramètre vectoriel* θ, et $E(\widehat{\theta})$ est alors le vecteur θ.

On définit le *risque quadratique scalaire* d'un estimateur θ^* de θ, noté $R(\theta^*, \theta)$, par :

$$R(\theta^*, \theta) = E\Big(\sum_{j=1}^{p} (\theta_j^* - \theta_j)^2\Big)$$

Dans cette partie, on cherche un estimateur θ^* de θ, représenté par un vecteur aléatoire $(\theta_1^*, \ldots, \theta_p^*)$, dont le risque $R(\theta^*, \theta)$ est strictement inférieur à $R(\widehat{\theta}, \theta)$.

1. Justifier que la variable aléatoire B_p suit la loi $\chi^2(p, b_p)$, et qu'elle constitue un estimateur biaisé de b_p.

2. On pose : $\theta^* = \left(1 - \frac{c}{B_p}\right) \times \widehat{\theta}$, où c est un paramètre réel strictement positif. Soit K une variable aléatoire qui suit la loi de Poisson de paramètre $\frac{b_p}{2}$.

a) En admettant que l'on a : $E\Big(\frac{1}{B_p} \sum_{j=1}^{p} \theta_j \, \widehat{\theta_j}\Big) = E\left(\frac{2K}{p-2+2K}\right)$, établir l'égalité suivante :

$$R(\theta^*, \theta) - R(\widehat{\theta}, \theta) = (c^2 - 2c(p-2)) \times E\left(\frac{1}{p-2+2K}\right)$$

b) Montrer que l'inégalité : $R(\theta^*, \theta) < R(\widehat{\theta}, \theta)$, est vérifiée si et seulement si : $0 < c < 2(p-2)$.
Déterminer en fonction de p, la valeur de c pour laquelle $R(\theta^*, \theta) - R(\widehat{\theta}, \theta)$ est minimale.
Comment s'écrit alors l'estimateur θ^* ?

Le sujet

Le problème de cette année avait pour objet la démonstration de « l'inadmissibilité » (risque quadratique non minimal), lorsque $p \geq 3$, de l'estimateur $\hat{\theta}$ du maximum de vraisemblance de l'espérance θ d'une loi normale à p dimensions, en mettant en évidence un estimateur $\bar{\theta}$ biaisé mais de risque quadratique inférieur à celui de $\hat{\theta}$ (estimateur de James-Stein).

Les principaux outils mathématiques utilisés étaient la transformation de Laplace, le théorème de transfert, la loi normale et ses dérivées (loi γ, loi du $\chi 2$), les lois conditionnelles, les notions d'estimateur et d'espérance conditionnelle.

Le préliminaire, faisant essentiellement appel à des techniques de calcul, se proposait de déterminer la transformée de Laplace de différentes lois de variables aléatoires. Les réponses à toutes les questions de ce préliminaire étaient données dans l'énoncé : on attendait donc des candidats une rédaction très précise et argumentée.

Dans la partie I, relativement classique, on étudiait quelques propriétés de la loi du χ^2 centré, et on établissait la formule d'un intervalle de confiance au risque α pour la variance σ^2 d'une variable aléatoire suivant une loi normale centrée réduite.

La partie II généralisait la partie précédente en introduisant la loi du χ^2 décentré à n degrés de liberté et en calculant la transformée de Laplace d'une telle loi.

La partie III, en étendant la formule de l'espérance totale au cas des variables aléatoires à densité, permettait de mettre en évidence la loi d'une somme de variables aléatoires.

Enfin dans la partie IV, on comparait les risques quadratiques des deux estimateurs $\bar{\theta}$ et $\hat{\bar{\theta}}$, ce qui permettait d'obtenir l'expression de l'estimateur de James-Stein.

Les résultats statistiques

Sur l'ensemble des 3114 candidats ayant composé dans cette épreuve, la note moyenne est de 9,88 avec un écart-type particulièrement élevé de 5,57 : il est vraisemblable que ce sujet a joué son rôle en classant les candidats tout en distinguant les meilleurs d'entre eux.
Les résultats par école sont :

- *HEC (2.154 candidats) – moyenne : 11,67 ; écart-type : 5,15.*
- *ESCP- EAP (2.571 candidats) – moyenne : 10,94 ; écart-type : 5,31.*

Erreurs les plus fréquentes

Les candidats, dans leur majorité, entreprennent beaucoup de calculs : ils n'utilisent pas les résultats des questions précédentes et font rarement appel au théorème du transfert. De plus, ils confondent souvent « équivalence » et « implication ».

Préliminaire (22% de la note finale).
On y relève de nombreuses erreurs liées à une maîtrise insuffisante du cours d'analyse, notamment des questions relatives à la convergence des intégrales impropres. Ainsi :

- *Certains candidats exhibent une primitive de exp $(-x^2)$.*
- *On voit très souvent le « théorème » suivant : une fonction f est intégrable sur $[0,+\infty[$ si et seulement si $\lim_{x\to\infty} x^2 f(x) = 0$.*
- *$\int_{-X}^{+X} f(x)\,dx$ tend vers 0 lorsque X tend vers $+\infty$, donc $\int_{-\infty}^{+\infty} f(x)\,dx$ existe et vaut 0.*
- *$\int_0^{+\infty} f(x)\,dx$ existe si et seulement si $f(x)$ tend vers 0 lorsque x tend vers $+\infty$.*
- *Au voisinage de $+\infty$, on a $ax^2 + bx \sim ax^2$, donc $\exp(ax^2 + bx) \sim \exp(ax^2)$*
- *Puisque $\int_{-\infty}^{+\infty} exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}$, on remplace 1/2 par a et on a : $\int_{-\infty}^{+\infty} exp(-ax^2)\,dx = \sqrt{\frac{\pi}{a}}$.*
- *Si $a \geq 0$, alors $\exp(-ax^2) \leq \exp(-x^2)$*

Partie I (21% de la note finale).

Dans la question 1, l'étude d'une fonction du niveau de classe Terminale et sa représentation graphique semblent relever de la haute virtuosité. Ainsi a-t-on vu des densités négatives, des confusions entre f' et f'', des points anguleux au sommet de la courbe. Une très faible proportion de candidats fait apparaître les tangentes à l'origine et la limite en $+\infty$ *dans l'étude des variations et le tracé.*

La question 4 relative à la détermination d'un intervalle de confiance, n'a été traitée que dans quelques copies ; la représentation graphique de la question 1 aurait dû permettre aux candidats de visualiser la non-symétrie de cet intervalle.

Partie II (20% de la note finale).

Cette partie fut peu traitée par les candidats ; parmi ceux qui l'ont abordée, on constate de nombreuses erreurs concernant la notion d'indépendance. Ainsi :

- $E(U^3) = 0$ *car* $E(U^3) = E(U^2)E(U)$, *puisque* U *et* U^2 *sont indépendantes, et également,* $E(U^4) = \left(E(U^2)\right)^2$.
- *L'indépendance des variables aléatoires n'est pas invoquée dans la stabilité de la loi gamma.*
- *Des variables aléatoires de même loi sont égales, donc* $W_n = nX_1$ *car les* X_i *suivent la même loi.*

Partie III (28% de la note finale).

Cette partie fait largement appel à la notion d'espérance conditionnelle et généralise la formule de l'espérance totale à une variable aléatoire à densité. La question 1 permet de vérifier cette formule sur un exemple, et la question 2, qui n'a pas été souvent abordée (en particulier la question 2c), avait pour finalité l'expression de l'espérance de l'inverse d'une variable aléatoire qui suit une loi du $\chi 2$ *décentré à l'aide de l'espérance d'une fonction d'une variable aléatoire de Poisson.*

Hormis les questions non traitées par les candidats, les principales erreurs proviennent des confusions entre variable aléatoire et loi de probabilité : on note toujours dans les copies, des « variables conditionnelles » ou des « événements conditionnels », et il est rare de rencontrer la définition d'une loi conditionnelle ou de la probabilité conditionnelle de A sachant B (pourtant au programme de la première année).

Partie IV (9% de la note finale).

De nombreux candidats ont entrepris directement la résolution de cette courte partie. Les calculs étaient élémentaires, mais même les dernières questions (calcul de c et écriture de $\bar{\theta}$) ont réservé bien des surprises...

Enfin pour terminer, signalons l'apparition dans certaines copies (heureusement peu nombreuses) d'expressions pour le moins obscures : « on procède par pilotage riemannien », « facettes de la fonction gamma » ou « d'après le CCIIFP ».

Recommandations aux futurs candidats

Le jury demande aux candidats une lecture attentive du texte qui précède toute épreuve de mathématiques, dans lequel il est précisé que la lisibilité et la qualité de la rédaction entrent pour une part non négligeable dans l'appréciation des copies.

Il est également conseillé de numéroter ses questions et d'encadrer ses résultats. Les raisonnements doivent être clairs et précis et les affirmations argumentées. Un apprentissage sérieux et une connaissance approfondie du cours sont donc indispensables.